# A multi-task learning approach to enhance sustainable biomolecule production in engineered microorganisms

**Erin H. Wilson** [1] **Mary E. Lidstrom** [2 3] **David A. C. Beck** [2 4]

## Abstract

A sustainable alternative to sourcing many materials humans need is metabolic engineering: a field that aims to engineer microorganisms into biological factories that convert renewable feedstocks into valuable biomolecules (i.e., jet fuel, medicine). In order for metabolic engineering to be cost-competitive, microorganism factories must be genetically optimized using predictable DNA sequence tools; however, for many organisms, the exact DNA sequence signals defining their genetic control systems are poorly understood. To better decipher these DNA signals, we propose a multi-task learning approach that uses deep learning and feature attribution methods to identify DNA sequence signals that control gene expression in the methanotroph *M. buryatense*. This bacterium consumes methane, a potent greenhouse gas. If successful, this work would enhance our ability to build gene expression tools to more effectively engineer *M. buryatense* into an efficient biomolecule factory that can divert methane pollution into valuable, everyday materials.

## 1. Introduction

Globally, human societies are consuming finite resources at unsustainable rates. Transitioning away from our dependencies on non-renewable resources and towards a cyclical, sustainable use of natural products is critical for reducing greenhouse gas emissions, preserving Earth's most threatened ecosystems, and securing longer term economic stability. Metabolic engineering is a growing field that aims to address sustainability concerns by engineering microorganisms into tiny biological factories that can convert renewable feedstocks (e.g., sugar cane) or waste streams (e.g., methane emissions, industrial waste-gas) into essential products like biofuels, medicines, and a wide range of biologically-derived materials (31, 28, 12). In order for microbial metabolic engineering to succeed as an alternative platform for producing valuable molecules that are otherwise sourced unsustainably, the process must be economically viable. From a biological engineering standpoint, this means that the microbial factories must be genetically optimized to produce the target molecules efficiently and at high yields (44).

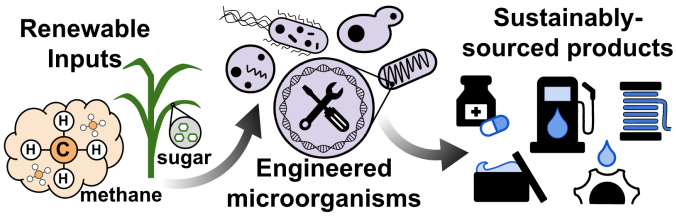


*Figure 1.* Metabolic engineering: a process to engineer microorganisms into biological factories that convert renewable inputs into more-sustainably sourced products.

Increasingly, automation and predictive modeling have been essential for accelerating the ability of metabolic engineering platforms to compete with industries rooted in petroleum, fossil fuels, or other unsustainable practices (27, 24). Many facets of metabolic engineering pipelines have improved with machine learning, such as metabolic network prediction (14, 30), bioreactor and fermentation process optimization (34, 6), and protein engineering (3, 19). Organism engineering – in particular reliably predicting biological outcomes from newly installed DNA parts – is another area of metabolic engineering that could benefit from machine learning. Organisms execute genetic programs using complex systems of signals that are encoded as DNA sequence patterns. The combination and orientation of these DNA patterns intricately regulate gene expression, or the timing and strength at which each gene turns on or off. While various approaches are being pursued (13, 46, 10), the precise rules of these "genetic grammars" are still poorly understood and the task of predicting gene expression output in engineered microorganisms remains difficult.

Deep learning methods, such as convolutional neural net-

[1]The Paul G. Allen School of Computer Science & Engineering, University of Washington [2]Department of Chemical Engineering, University of Washington [3]Department of Microbiology, University of Washington [4]eScience Institute, University of Washington. Correspondence to: Erin Wilson <ewilson6@uw.edu>.

*Tackling Climate Change with Machine Learning workshop at ICML 2021.*

works (CNNs) and recurrent neural networks (RNNs), are well-suited to DNA sequence pattern discovery tasks: they are particularly adept at learning important features without prior knowledge, finding relevant patterns within larger contexts, and considering non-linear or longer-term dependencies between learned features (22, 16). We propose using a multi-task deep learning approach to elucidate genetic grammar rules in microorganisms with potential to serve as metabolic engineering platforms. Specifically, by 1) using deep learning model architectures to predict gene expression strength across a variety of growth conditions directly from DNA sequences and 2) applying feature attribution methods to identify meaningful patterns within the DNA inputs, we can use these discovered patterns to develop genetic tools required to optimize microbes to produce valuable molecules efficiently, sustainably, and at large scales.

## 2. Background

### 2.1. Sustainable Biomolecule Production

Humans rely on many biologically-derived molecules: fuels for transportation, fibers in clothing, medicinal molecules from plants. Molecules naturally found in organisms are typically produced via some metabolic pathway, or a series of chemical conversions carried out by enzymes that can transform inputs, like sugars, into other molecules organisms need to survive. Organisms store instructions for building metabolic pathway enzymes in DNA. Since DNA is a common language between all organisms, genetic instructions are potentially transferable between species. Metabolic engineers leverage this genetic transferability to rewire metabolic pathways in microorganisms, like bacteria, to produce a range of valuable molecules that other organisms, like plants, make naturally (31, 28, 12).

One of the earliest successful examples was an effort to re-engineer baker's yeast to convert sugarcane into artemisinin (35), a key component in malaria treatments originally found in the sweet wormwood plant. Since then, many other molecules, such as farnesene (26) (a jet fuel) and spider-silk (42), have similarly been produced in microbes. These examples demonstrate the ability of metabolic engineering strategies to support sustainable biomolecule production, but microbes' production efficiency must continue to improve in order to be economically competitive.

### 2.2. Genetic Challenges in Metabolic Engineering

A major challenge of microbial optimization is that each gene in a newly installed metabolic pathway must have finely-tuned expression. Organisms have evolved intricate systems of controls to regulate gene expression, namely genetic signals encoded as DNA sequence patterns. These sequence patterns, or motifs, exist throughout the genome and are often short and can be arranged in many different combinations and orientations (7, 21, 8). Cells understand these motif patterns as a "genetic grammar" and use them to perform logical operations to determine which genes need to be activated or repressed in response to the current environmental conditions. Promoters are regions of DNA that contain many of the sequence motifs involved in gene regulation. Therefore promoter regions are key elements to identify and decode, both for better elucidating a microbe's basic biology as well as for building out a genetic toolkit with which to more precisely and effectively engineer the microbe for biomolecule production (4).

While many regulatory signals, such as promoters, have been identified and studied in popular microorganisms like baker's yeast and *E. coli*, there are countless other microbial species that have not yet had the same degree of genetic characterization. Leveraging the diversity of microbes across the tree of life, many of which could serve as ideal platforms for metabolic engineering, would broaden the opportunities for this renewable production strategy to succeed. Unfortunately, every organism has evolved a distinct genetic grammar and though some may be similar, promoter tools developed for one organism are not always compatible across species (32, 43). If we could accelerate our ability to develop the necessary tools with which to engineer less-studied microorganisms, it would greatly enhance the potential for metabolic engineering to become an economically viable molecule production strategy by reducing the time and investment needed to rapidly explore new potential host organisms.

### 2.3. Methane Emissions Mitigation

One promising microbial host is the methanotroph *Methylotuvimicrobium buryatense* 5GB1, a bacterium that can use one-carbon compounds, such as methane and methanol, to grow and survive (18, 11). Methane is emitted from both natural (e.g., wetlands) and anthropogenic sources (e.g., landfills, coal mines, agriculture) and is the second greatest contributor to climate change behind carbon dioxide (33). Though less abundant than carbon dioxide, methane is 20-30x more potent as a greenhouse gas and thus addressing methane emissions is a critical avenue for mitigating climate impacts (38).

Methanotrophs like *M. buryatense* play important roles in consuming methane and cycling carbon back into the environment (11). Methane concentrations tend to be enriched in the atmosphere surrounding industrial sites that emit the gas as a byproduct and thus there is an opportunity to mitigate emissions at these types of pollution sources using bioreactors designed for growing methanotrophs (23). In particular, if an efficient metabolic engineering system could be deployed with a methane-consuming microbe, it could

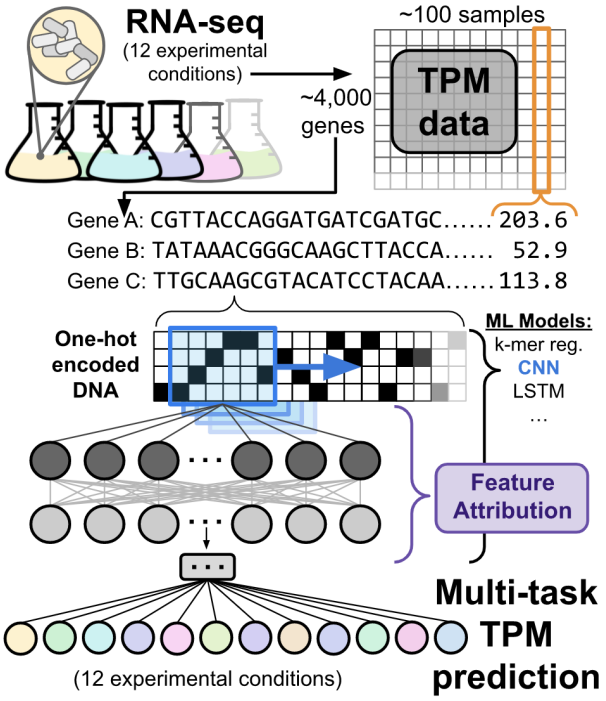


*Figure 2.* Multi-task learning approach. RNA-seq data measuring transcripts per million (TPM) were collected for ∼4,000 genes in ∼100 samples. Each sample belongs to one of 12 experimental growth conditions. One-hot encoded upstream DNA sequences will be fed into varying model architectures (linear regression, CNN, LSTM) to predict genes' TPM output in a multi-task framework. Feature attribution methods will be applied to identify influential sequence motifs.

offer an attractive outlet for methane emissions: a feedstock for biological factories. Not only would this provide another paradigm in which to harness biology for sustainable molecule production, but it would help divert a harmful waste stream out of the atmosphere and sequester it in useful materials. Continued innovations in methane capture and bioreactor technologies are required in order to scale up this intervention, however the ability to develop an engineered methanotroph with optimized metabolism is a crucial step and the primary focus of this proposal.

## 3. Technical Approach

### 3.1. Related Work

To develop more sophisticated genetic tools for efficiently engineering *M. buryatense*, we aim to use a deep learning approach to learn DNA sequence patterns from its promoter regions. Deep learning has previously been applied to DNA sequence inputs, for example to predict the presence of DNA regulatory sites (2, 45, 20), estimate strength from a sequence (13, 36, 5), or classify sequences as promoters (40, 29). However most of these approaches tend to focus on model organisms, like human, mouse, yeast, or *E. coli*, with vast amounts of experimental data. *M. buryatense*, and many other non-model bacteria, do not have databases (37, 15) of such extensively curated knowledge. Deep learning approaches that can leverage simple, routine-to-collect datasets to learn relevant signalling patterns in unusual organisms would enable more rapid development of genetic tools for diverse species.

RNA-sequencing is a common experimental technique used to measure transcription, an important aspect of gene expression strength (41). Briefly, it takes a snapshot of the RNA transcript levels of every gene in the cell, revealing which genes the cell has currently received signals to activate and their approximate expression strength. RNA transcript abundances change in response to these signals, producing a valuable readout with which to interrogate signalling patterns in promoter regions with predictive models (1, 46).

### 3.2. Dataset and Multi-task Learning

We have compiled an RNA-seq dataset recording the expression strength in transcripts per million (TPM) of each of the ∼4,000 genes in the *M. buryatense* genome. Each gene was repeatedly measured in ∼100 experimental samples. Each sample is labeled with one of 12 possible experimental growth conditions (e.g., "ideal conditions", "methane limited", "no copper"). Additionally, from the *M. buryatense* gene annotation file (17), we have extracted the upstream DNA sequences of each gene, a region likely to contain promoters and other regulatory signals.

Using one-hot encoded DNA sequences as input, we will apply a suite of machine learning architectures to predict the TPM levels for each gene in each condition. Specifically, we plan to compare simpler models, such as linear regression on k-mer counts, to more complex deep learning architectures, such as CNNs and LSTMs, and evaluate regression losses using the Mean Squared Error. Given that some experimental growth conditions are more similar than others, we will use a multi-task framework to simultaneously estimate TPM in each experimental growth condition, allowing the model to share learned features that are relevant across multiple prediction tasks. Furthermore, we intend to apply feature attribution methods, such as DeepLift (39), DeepShap (9), and Scrambler Networks (25) to identify meaningful subsequences within the inputs that influence expression in specific conditions. These subsequences are likely to represent regulatory motifs that can form the basis for new genetic engineering tools for this organism.

While our RNA-seq dataset is unique in its diversity of experimental conditions for such an unusual organism, *M. buryatense's* genome of 4,000 genes is small relative to the wider set of microbe gene expression data available. We anticipate that using transfer learning techniques to pre-train models using data from related tasks or related organisms

will be quite valuable, enabling us to learn more universally conserved signalling patterns from a larger dataset before fine-tuning models to learn the specifics of *M. buryatense's* genetic grammar.

## 4. Conclusions and Impacts

If successful, this work would enable us to 1) predict the influence of new DNA sequences on *M. buryatense* gene expression in a range of conditions, estimating their effectiveness as candidate promoters, 2) gain biological insights about specific sequence motifs that emerge from model features flagged as particularly important for making predictions, as most regulatory features are not currently known for this organism, and 3) use discovered sequence motifs to build DNA parts, like synthetic promoters, to more effectively control foreign genes in newly installed metabolic pathways. Deep learning approaches have already seen successes in model organisms with plenty of data. This proposal explicitly aims to extend these approaches to non-model organisms that have thus far received less experimental attention but still warrant genetic characterization due to their potential to serve as metabolic engineering platforms.

*M. buryatense* is a promising microbe that, if effectively engineered, could divert methane emissions into useful products. This approach may similarly be applied to other organisms with desirable metabolic properties, enhancing our ability to develop genetic tools more broadly. Overall, we aim to extend the reach of machine learning to metabolic engineering in non-model organisms, a field with direct avenues for impacting climate change by enabling sustainable molecule production and redirecting harmful emissions into valuable materials.

## 5. Acknowledgments

Many thanks to Dr. Nathan Hodas for his advice and mentorship throughout the proposal writing process. Additionally, E.H.W. acknowledges support from the NSF GRFP (DGE-1762114). M.E.L. acknowledges support from the University of Washington. D.A.C.B. acknowledges support from HDR: I-DIRSE-FW: Accelerating the Engineering Design and Manufacturing Life-Cycle with Data Science under Grant No. NSF 1934292.

## References

[1] AGARWAL, V., AND SHENDURE, J. Predicting mRNA Abundance Directly from Genomic Sequence Using Deep Convolutional Neural Networks. *Cell Reports 31*, 7 (May 2020), 107663.

[2] ALIPANAHI, B., DELONG, A., WEIRAUCH, M. T., AND FREY, B. J. Predicting the sequence specificities of DNA- and RNA-binding proteins by deep learning. *Nature Biotechnology 33*, 8 (Aug. 2015), 831–838.

[3] ALLEY, E. C., KHIMULYA, G., BISWAS, S., ALQURAISHI, M., AND CHURCH, G. M. Unified rational protein engineering with sequence-based deep representation learning. *Nature Methods 16*, 12 (Dec. 2019), 1315–1322. Number: 12 Publisher: Nature Publishing Group.

[4] BERVOETS, I., AND CHARLIER, D. Diversity, versatility and complexity of bacterial gene regulation mechanisms: opportunities and drawbacks for applications in synthetic biology. *FEMS Microbiology Reviews* (Feb. 2019).

[5] BOGARD, N., LINDER, J., ROSENBERG, A. B., AND SEELIG, G. A Deep Neural Network for Predicting and Engineering Alternative Polyadenylation. *Cell 0*, 0 (June 2019).

[6] BRADFORD, E., SCHWEIDTMANN, A. M., ZHANG, D., JING, K., AND DEL RIO-CHANONA, E. A. Dynamic modeling and optimization of sustainable algal production with uncertainty using multivariate Gaussian processes. *Computers & Chemical Engineering 118* (Oct. 2018), 143–158.

[7] BROWNING, D. F., AND BUSBY, S. J. W. The regulation of bacterial transcription initiation. *Nature Reviews Microbiology 2*, 1 (Jan. 2004), 57–65. Number: 1 Publisher: Nature Publishing Group.

[8] BROWNING, D. F., BUTALA, M., AND BUSBY, S. J. W. Bacterial Transcription Factors: Regulation by Pick "N" Mix. *Journal of Molecular Biology 431*, 20 (Sept. 2019), 4067–4077.

[9] CHEN, H., LUNDBERG, S., AND LEE, S.-I. Explaining Models by Propagating Shapley Values of Local Components. *arXiv:1911.11888 [cs, stat]* (Nov. 2019). arXiv: 1911.11888.

[10] CHEN, L., AND CAPRA, J. A. Learning and interpreting the gene regulatory grammar in a deep learning framework. *PLOS Computational Biology 16*, 11 (Nov. 2020), e1008334. Publisher: Public Library of Science.

[11] CHISTOSERDOVA, L. Modularity of methylotrophy, revisited. *Environmental Microbiology 13*, 10 (2011), 2603–2622. _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1111/j.1462-2920.2011.02464.x.

[12] CRAVENS, A., PAYNE, J., AND SMOLKE, C. D. Synthetic biology strategies for microbial biosynthesis of plant natural products. *Nature Communications 10*,

1 (May 2019), 2142. Number: 1 Publisher: Nature Publishing Group.

[13] CUPERUS, J. T., GROVES, B., KUCHINA, A., ROSENBERG, A. B., JOJIC, N., FIELDS, S., AND SEELIG, G. Deep learning of the regulatory grammar of yeast 5' untranslated regions from 500,000 random sequences. *Genome Research 27*, 12 (Dec. 2017), 2015–2024.

[14] DALE, J. M., POPESCU, L., AND KARP, P. D. Machine learning methods for metabolic pathway prediction. *BMC Bioinformatics 11*, 1 (Jan. 2010), 15.

[15] FORNES, O., CASTRO-MONDRAGON, J. A., KHAN, A., VAN DER LEE, R., ZHANG, X., RICHMOND, P. A., MODI, B. P., CORREARD, S., GHEORGHE, M., BARANAŠIĆ, D., SANTANA-GARCIA, W., TAN, G., CHÈNEBY, J., BALLESTER, B., PARCY, F., SANDELIN, A., LENHARD, B., WASSERMAN, W. W., AND MATHELIER, A. JASPAR 2020: update of the open-access database of transcription factor binding profiles. *Nucleic Acids Research 48*, D1 (Jan. 2020), D87–D92.

[16] GRAVES, A., AND SCHMIDHUBER, J. Framewise phoneme classification with bidirectional LSTM and other neural network architectures. *Neural Networks 18*, 5 (July 2005), 602–610.

[17] GROOM, J., FORD, S., PESESKY, M., AND LIDSTROM, M. Methylomicrobium buryatense strain 5GB1C chromosome, complete genome. *NCBI Nucleotide Database* (June 2019). {:itemType: dataset}.

[18] HANSON, R. S., AND HANSON, T. E. Methanotrophic bacteria. *Microbiological Reviews 60*, 2 (June 1996), 439–471.

[19] HIRANUMA, N., PARK, H., BAEK, M., ANISHCHENKO, I., DAUPARAS, J., AND BAKER, D. Improved protein structure refinement guided by deep learning based accuracy estimation. *Nature Communications 12*, 1 (Feb. 2021), 1340. Number: 1 Publisher: Nature Publishing Group.

[20] KELLEY, D. R., SNOEK, J., AND RINN, J. L. Basset: learning the regulatory code of the accessible genome with deep convolutional neural networks. *Genome Research 26*, 7 (July 2016), 990–999.

[21] KOSURI, S., GOODMAN, D. B., CAMBRAY, G., MUTALIK, V. K., GAO, Y., ARKIN, A. P., ENDY, D., AND CHURCH, G. M. Composability of regulatory sequences controlling transcription and translation in Escherichia coli. *Proceedings of the National Academy of Sciences 110*, 34 (Aug. 2013), 14024–14029.

[22] KRIZHEVSKY, A., SUTSKEVER, I., AND HINTON, G. ImageNet Classification with Deep Convolutional Neural Networks. *Neural Information Processing Systems 25* (Jan. 2012).

[23] LA, H., HETTIARATCHI, J. P. A., ACHARI, G., AND DUNFIELD, P. F. Biofiltration of methane. *Bioresource Technology 268* (Nov. 2018), 759–772.

[24] LAWSON, C. E., MARTÍ, J. M., RADIVOJEVIC, T., JONNALAGADDA, S. V. R., GENTZ, R., HILLSON, N. J., PEISERT, S., KIM, J., SIMMONS, B. A., PETZOLD, C. J., SINGER, S. W., MUKHOPADHYAY, A., TANJORE, D., DUNN, J. G., AND GARCIA MARTIN, H. Machine learning for metabolic engineering: A review. *Metabolic Engineering 63* (Jan. 2021), 34–60.

[25] LINDER, J., FLEUR, A. L., CHEN, Z., LJUBETIČ, A., BAKER, D., KANNAN, S., AND SEELIG, G. Interpreting Neural Networks for Biological Sequences by Learning Stochastic Masks. *bioRxiv* (Apr. 2021), 2021.04.29.441979. Publisher: Cold Spring Harbor Laboratory Section: New Results.

[26] MEADOWS, A. L., HAWKINS, K. M., TSEGAYE, Y., ANTIPOV, E., KIM, Y., RAETZ, L., DAHL, R. H., TAI, A., MAHATDEJKUL-MEADOWS, T., XU, L., ZHAO, L., DASIKA, M. S., MURARKA, A., LENIHAN, J., ENG, D., LENG, J. S., LIU, C.-L., WENGER, J. W., JIANG, H., CHAO, L., WESTFALL, P., LAI, J., GANESAN, S., JACKSON, P., MANS, R., PLATT, D., REEVES, C. D., SAIJA, P. R., WICHMANN, G., HOLMES, V. F., BENJAMIN, K., HILL, P. W., GARDNER, T. S., AND TSONG, A. E. Rewriting yeast central carbon metabolism for industrial isoprenoid production. *Nature 537*, 7622 (Sept. 2016), 694–697.

[27] MOWBRAY, M., SAVAGE, T., WU, C., SONG, Z., CHO, B. A., DEL RIO-CHANONA, E. A., AND ZHANG, D. Machine learning for biochemical engineering: A review. *Biochemical Engineering Journal 172* (Aug. 2021), 108054.

[28] NIELSEN, J., AND KEASLING, J. D. Engineering Cellular Metabolism. *Cell 164*, 6 (Mar. 2016), 1185–1197.

[29] OUBOUNYT, M., LOUADI, Z., TAYARA, H., AND CHONG, K. T. DeePromoter: Robust Promoter Predictor Using Deep Learning. *Frontiers in Genetics 10* (Apr. 2019).

[30] OYETUNDE, T., ZHANG, M., CHEN, Y., TANG, Y., AND LO, C. BoostGAPFILL: improving the fidelity of metabolic network reconstructions through integrated constraint and pattern-based methods. *Bioinformatics (Oxford, England) 33*, 4 (Feb. 2017), 608–611.

[31] PICKENS, L. B., TANG, Y., AND CHOOI, Y.-H. Metabolic Engineering for the Production of Natural Products. *Annual review of chemical and biomolecular engineering 2* (2011), 211–236.

[32] PORTELA, R. M. C., VOGL, T., KNIELY, C., FISCHER, J. E., OLIVEIRA, R., AND GLIEDER, A. Synthetic Core Promoters as Universal Parts for Fine-Tuning Expression in Different Yeast Species. *ACS Synthetic Biology 6*, 3 (Mar. 2017), 471–484. Publisher: American Chemical Society.

[33] PRATT, C., AND TATE, K. Mitigating Methane: Emerging Technologies To Combat Climate Change's Second Leading Contributor. *Environmental Science & Technology 52*, 11 (June 2018), 6084–6097.

[34] PSICHOGIOS, D. C., AND UNGAR, L. H. A hybrid neural network-first principles approach to process modeling. *AIChE Journal 38*, 10 (Oct. 1992), 1499–1511.

[35] RO, D.-K., PARADISE, E. M., OUELLET, M., FISHER, K. J., NEWMAN, K. L., NDUNGU, J. M., HO, K. A., EACHUS, R. A., HAM, T. S., KIRBY, J., CHANG, M. C. Y., WITHERS, S. T., SHIBA, Y., SARPONG, R., AND KEASLING, J. D. Production of the antimalarial drug precursor artemisinic acid in engineered yeast. *Nature 440*, 7086 (Apr. 2006), 940–943.

[36] SAMPLE, P. J., WANG, B., REID, D. W., PRESNYAK, V., MCFADYEN, I., MORRIS, D. R., AND SEELIG, G. Human 5' UTR design and variant effect prediction from a massively parallel translation assay. *Nature biotechnology 37*, 7 (July 2019), 803–809.

[37] SANTOS-ZAVALETA, A., SALGADO, H., GAMA-CASTRO, S., SÁNCHEZ-PÉREZ, M., GÓMEZ-ROMERO, L., LEDEZMA-TEJEIDA, D., GARCÍA-SOTELO, J. S., ALQUICIRA-HERNÁNDEZ, K., MUÑIZ-RASCADO, L. J., PEÑA-LOREDO, P., ISHIDA-GUTIÉRREZ, C., VELÁZQUEZ-RAMÍREZ, D. A., DEL MORAL-CHÁVEZ, V., BONAVIDES-MARTÍNEZ, C., MÉNDEZ-CRUZ, C.-F., GALAGAN, J., AND COLLADO-VIDES, J. RegulonDB v 10.5: tackling challenges to unify classic and high throughput knowledge of gene regulation in E. coli K-12. *Nucleic Acids Research 47*, D1 (Jan. 2019), D212–D220.

[38] SAUNOIS, M., STAVERT, A. R., POULTER, B., BOUSQUET, P., CANADELL, J. G., JACKSON, R. B., RAYMOND, P. A., DLUGOKENCKY, E. J., HOUWELING, S., PATRA, P. K., CIAIS, P., ARORA, V. K., BASTVIKEN, D., BERGAMASCHI, P., BLAKE, D. R., BRAILSFORD, G., BRUHWILER, L., CARLSON, K. M., CARROL, M., CASTALDI, S., CHANDRA, N., CREVOISIER, C., CRILL, P. M., COVEY, K., CURRY, C. L., ETIOPE, G., FRANKENBERG, C., GEDNEY, N., HEGGLIN, M. I., HÖGLUND-ISAKSSON, L., HUGELIUS, G., ISHIZAWA, M., ITO, A., JANSSENS-MAENHOUT, G., JENSEN, K. M., JOOS, F., KLEINEN, T., KRUMMEL, P. B., LANGENFELDS, R. L., LARUELLE, G. G., LIU, L., MACHIDA, T., MAKSYUTOV, S., MCDONALD, K. C., MCNORTON, J., MILLER, P. A., MELTON, J. R., MORINO, I., MÜLLER, J., MURGUIA-FLORES, F., NAIK, V., NIWA, Y., NOCE, S., O'DOHERTY, S., PARKER, R. J., PENG, C., PENG, S., PETERS, G. P., PRIGENT, C., PRINN, R., RAMONET, M., REGNIER, P., RILEY, W. J., ROSENTRETER, J. A., SEGERS, A., SIMPSON, I. J., SHI, H., SMITH, S. J., STEELE, L. P., THORNTON, B. F., TIAN, H., TOHJIMA, Y., TUBIELLO, F. N., TSURUTA, A., VIOVY, N., VOULGARAKIS, A., WEBER, T. S., VAN WEELE, M., VAN DER WERF, G. R., WEISS, R. F., WORTHY, D., WUNCH, D., YIN, Y., YOSHIDA, Y., ZHANG, W., ZHANG, Z., ZHAO, Y., ZHENG, B., ZHU, Q., ZHU, Q., AND ZHUANG, Q. The Global Methane Budget 2000–2017. *Earth System Science Data 12*, 3 (July 2020), 1561–1623. Publisher: Copernicus GmbH.

[39] SHRIKUMAR, A., GREENSIDE, P., AND KUNDAJE, A. Learning Important Features Through Propagating Activation Differences. *arXiv:1704.02685 [cs]* (Oct. 2019). arXiv: 1704.02685.

[40] UMAROV, R. K., AND SOLOVYEV, V. V. Recognition of prokaryotic and eukaryotic promoters using convolutional deep learning neural networks. *PLOS ONE 12*, 2 (Feb. 2017), e0171410.

[41] WANG, Z., GERSTEIN, M., AND SNYDER, M. RNA-Seq: a revolutionary tool for transcriptomics. *Nature Reviews Genetics 10*, 1 (Jan. 2009), 57–63. Number: 1 Publisher: Nature Publishing Group.

[42] WIDMAIER, D. M., TULLMAN-ERCEK, D., MIRSKY, E. A., HILL, R., GOVINDARAJAN, S., MINSHULL, J., AND VOIGT, C. A. Engineering the Salmonella type III secretion system to export spider silk monomers. *Molecular Systems Biology 5*, 1 (Jan. 2009), 309.

[43] WILSON, E. H., GROOM, J. D., SARFATIS, M. C., FORD, S. M., LIDSTROM, M. E., AND BECK, D. A. C. A Computational Framework for Identifying Promoter Sequences in Nonmodel Organisms Using RNA-seq Data Sets. *ACS Synthetic Biology* (May 2021). Publisher: American Chemical Society.

[44] WOOLSTON, B. M., EDGAR, S., AND STEPHANOPOULOS, G. Metabolic Engineering: Past and Future. *Annual Review of Chemical and*

*Biomolecular Engineering 4*, 1 (June 2013), 259–288. Publisher: Annual Reviews.

[45] ZHOU, J., AND TROYANSKAYA, O. G. Predicting effects of noncoding variants with deep learning-based sequence model. *Nature Methods 12*, 10 (Oct. 2015), 931–934.

[46] ZRIMEC, J., BÖRLIN, C. S., BURIC, F., MUHAMMAD, A. S., CHEN, R., SIEWERS, V., VERENDEL, V., NIELSEN, J., TÖPEL, M., AND ZELEZNIAK, A. Deep learning suggests that gene expression is encoded in all parts of a co-evolving interacting gene regulatory structure. *Nature Communications 11*, 1 (Dec. 2020), 6141. Number: 1 Publisher: Nature Publishing Group.